

2012 NCES MIS Conference

Innovative, Relevant Longitudinal Data Warehousing/Reporting

Reusing Metrics as Attributes

Data modeling techniques to add a student
centric focus to a data warehouse

Agenda

- History of LCSD #1 Data Warehouse Initiative
 - How did we get here?
- Data Warehouse Fundamentals Review
- Introduction of Metric Driven Attributes
- Implementation Challenges and Design of Solutions
- Technical Implementation
- Summary and Q/A

About the speakers – Kyle McKinney

- Program Administrator, Information Systems
- Oversees district data system operations including PowerSchool, Pinnacle Gradebook, SAP and data warehouse/BI programs
- Nine years experience with district

About the speakers – Alan Moore, Ph.D.

- Program Administrator, Data Research Analyst
- Taught 24 years in educational statistics and measurement
- Former Director of Standards and Assessment Division for the Wyoming Department of Education
- Really good at chess

About the speakers – Trev Swarm

- Senior Systems Analyst
- Data warehouse architect
- BS in Computer Science
- 12 years experience in application development, database design and system administration
- Annoyingly bad at chess

Laramie County School District #1

- 13,200 students, 2,400 staff, covers 1,592 Sq. Miles, 34 campuses, 7 support sites.
- 13,000 network connected devices, a MOE WAN, 9,000 computers and 150 servers - many virtualized.
- We run SAP, PowerSchool, Pinnacle, Welligent, Trapeze, SDMS, with about 40 other district wide data silos, and hundreds of smaller scope data silos.

Why build a Data Warehouse?

- LCSD1 Strategic Plan 2010-2013
 - Focus Areas
 - Students will:
 - Achieve academic success
 - Graduate from high school
 - Be prepared for college and career
 - Be responsible citizens
 - Increase effective and efficient practices, and make better use of all of our resources
- Our current systems weren't doing what we wanted

Operationalizing the Focus Areas

- The district has become serious about operationalizing measures related to these focus areas
- It has also set out to track trends across time and to disaggregate data across schools
- So, examples of the kinds of measures that become important are:
 - Student standardized test scores
 - GPA
 - Honor Course taking patterns
 - Graduation rates
 - Dropout rates
 - Student scores on college entrance exams
 - Attendance patterns
 - School discipline patterns

Types of Questions

- The types of questions that are being asked are questions like:
 - What has been the pattern of student achievement on standardized tests like the state test and the interim assessment given 3 times per year?
 - How do the patterns of achievement differ across schools?
 - How do they differ by subgroup of students such as racial/ethnic groups or lunch status?
 - What is the trend in graduation rate?
 - Who is not graduating and why?
 - What is the effect of specific remedial programs on student learning, as measured by test scores?
 - What are the individual student profiles of test score performance over time?
 - How do strategies of different schools affect their attendance, school discipline problems and drop-out rates?

Challenges

- Lack of data validation on entry
- Limited longitudinal data
- Difficult to find some data
- Data doesn't follow students transferring to other schools in the district
- Complicated calculations for proficiencies
- Difficult to track standards students need for graduation

Data Silos

- "Inch Deep" or "Mile Wide"
- Difficult to gather valid & reliable data
- Data Silo Synchronization
- Resources required for silos to be real time
- Playing nice with others
- Service gaps between silos

Vision

- Answering these questions and meeting these challenges requires that the district data warehouse:
 - Support inquiry of student attributes over time.
 - Provide an easy to use resource for clean, easily accessible data, from different sources, for data driven decision-making.
 - Integrate disparate data sources to provide a single repository point and single version of “the truth.”

What We Want...

- Data about students, teachers, parents, staff, schools, district operations, and the community, from multiple sources
- Data over time and at points in time
- Prepared role based views of data: district, triad, level of school, school, course, teacher, and student

...that's not all

- The look & feel of our current district tools
 - MS Word, Excel, and PowerPoint
- Point & click use
- Self Service – Reports designed by district staff
- Drill down to granular data
- Switch report types – tables, graphs and charts
- Scalability, adaptability, and sustainability

Wouldn't it be great to have...!

- Collaborative workspaces (PLC district)
- Access from work or home
- Electronic portfolio of student work K-12
- Written, audio, video, work tied to grades / standards
- Track interventions, interventionists, and results
- Determine potential interventions for other students
- Determine training for interventionists
- Tie into potential state data warehouse
- Track students pre & post K-12 experience

Where is this tool?

- Looked for “the answer”:
 - Investigated published research projects
 - Investigated use of current systems
 - Investigated “off the shelf” systems
 - Investigated “best of breed” solutions found from the USDoE, State DoEs, higher education data systems, and other K12 districts

If you build it...

- Custom developed solution on Microsoft BI Stack
 - SQL Server – Analysis Services, Reporting Services, Integration Services
 - SharePoint 2010 Portal
 - PowerPivot

Open Menu



LARAMIE COUNTY SCHOOL DISTRICT #1

Search this site...



Provide Feedback...


Home

Schools

PLC

Departments ▾

District ▾



STRATEGIC PLAN 2010 2013

FOCUS AREAS

STUDENT ACHIEVEMENT

- Focus, implement and monitor achievement strategy for quality, equity, learning & excellence K-12
- Measure & guarantee & update curriculum to all students across all students
- Develop a comprehensive staff development plan that supports the strategic plan
- Develop achievement strategies, standards that are growth model of excellence

DATA-DRIVEN DECISION MAKING

- Identify data that will be consistently monitored, reported & used with fidelity, integrity & consistency
- Develop a culture of data-driven decisions making across the district
- Develop a process to gather, analyze & use data that is a ready resource for the classroom in the classroom
- Develop a process to monitor & update annual goals for success, equity, integrity & consistency

LEADERSHIP DEVELOPMENT

- Focus all leadership staff to successfully implement research-based leadership practices
- Develop a plan to provide leadership training to all staff
- Continue to build K-12 leadership capacity within and across the district
- Build student leadership skills and empower them to create responsible citizens

STRATEGIC COMMUNICATION


- Refine the process of communication for the governing board, administration, instructional support staff, parents & general community
- Expand communication strategies to utilize new technologies
- Develop internal & external communication that supports the focus areas

EFFECTIVE & EFFICIENT USE OF RESOURCES

- Improve efficiency & effectiveness of resources across the district, every person

THE MISSION OF LCSD 1:
In cooperation with parents, students, staff and the community to be a premier high quality education in a safe and secure environment for all students, inspiring them to become lifelong learners and responsible citizens.

STUDENTS WILL:
achieve academic success | graduate from high school
be prepared for college & careers | be responsible citizens



From the Superintendent



The Superintendent of Schools is the District's executive school officer. He is accountable to the Board of Trustees and is responsible for administering the Board's policies and general management of the District's operation. [\[More\]](#)

Mission Statement ▾

Vision Statement ▾

About Us ▾

School Board Approves 2010-2013 strategic plan [Read](#)

Parents

Students

Employees

Community

District Dashboards

Libraries

Lists

Discussions

Student Attendance - by Triad : [Daily Absences](#) | [Total Absences](#) | [High Absenteeism](#) | [Consecutive Absences](#) | [Dropout Rates](#) | [Side by Side Charts](#) | **Absence Codes**

Race: **All**

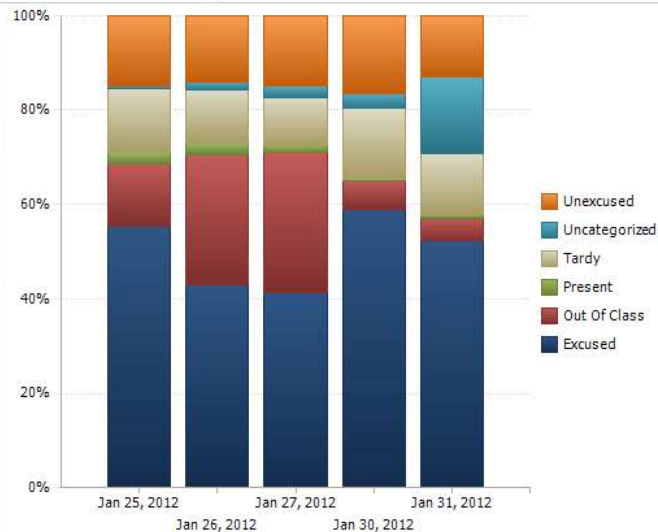
Hispanic: **All**

Triad: **District**

Lunch Status: **All**

IEP Status: **All**

Daily Absence Codes by Triad

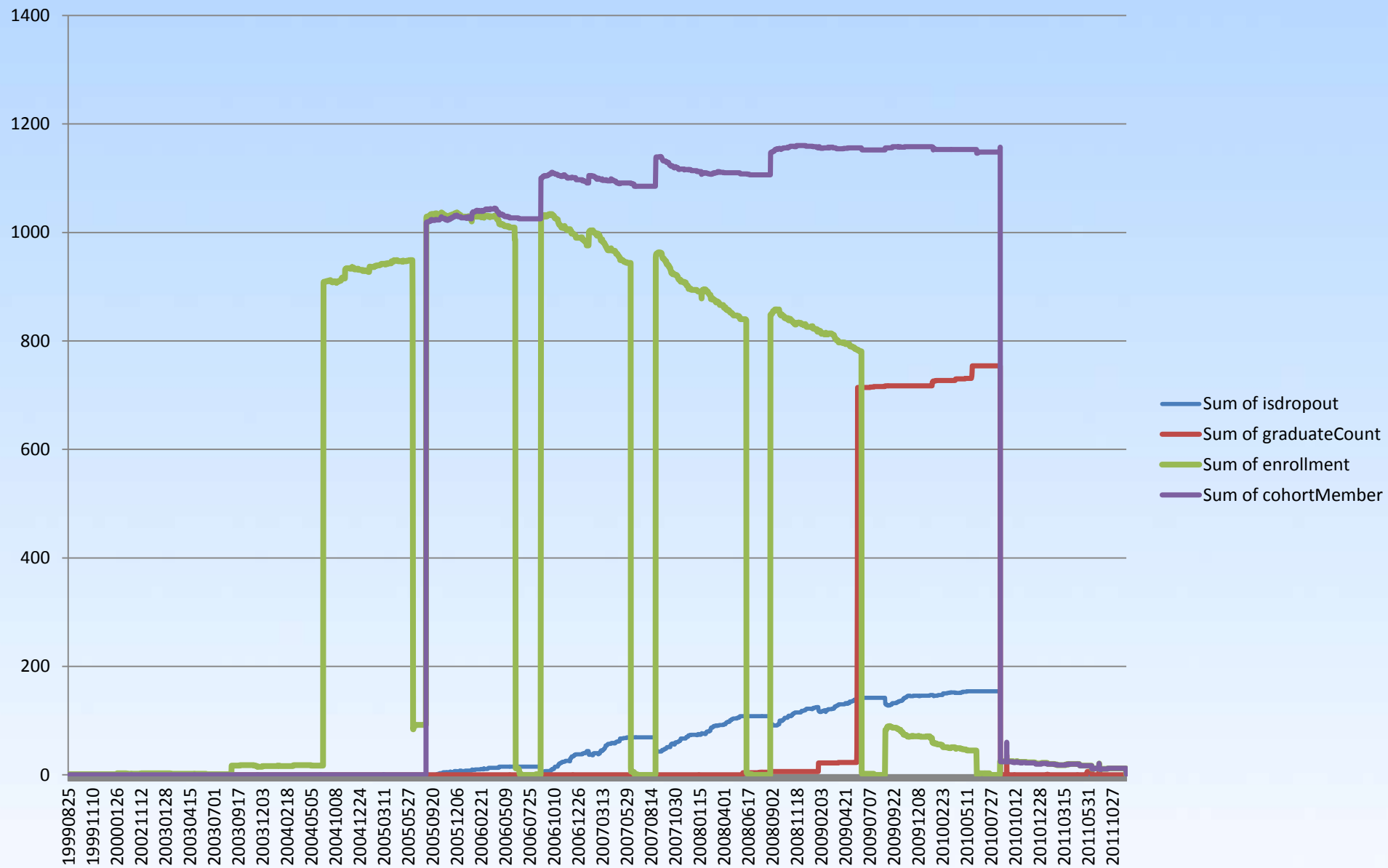


Daily Absence Codes Grid by Triad

School Triads, Class...	Jan 25, 2012	Jan 26, 2012	Jan 27, 2012	Jan 30, 2012	Jan 31, 2012
<input checked="" type="checkbox"/> District					
<input checked="" type="checkbox"/> Excused	1,418.00	1,340.00	2,775.00	1,656.00	1,318.00
<input checked="" type="checkbox"/> Out Of Class	334.00	852.00	1,998.00	176.00	114.00
<input checked="" type="checkbox"/> Present	66.00	82.00	90.00	9.00	19.00
<input checked="" type="checkbox"/> Tardy	346.00	349.00	687.00	420.00	324.00
<input checked="" type="checkbox"/> Uncategorized	15.00	54.00	157.00	90.00	414.00
<input checked="" type="checkbox"/> Unexcused	381.00	436.00	1,006.00	464.00	324.00

...and that's just the beginning...

		District 912 Contiguous ▼		
GPA Bucket ▼	Has Less Than Five Daily Absences YTD ▼	No	Yes	Grand Total
		Membership Value	Membership Value	Membership Value
☐ .99-0.00	No	14	29	43
	Yes	2	3	5
	Total	16	32	48
☐ 1.99-1.00	No	106	304	410
	Yes	54	115	169
	Total	160	419	579
☐ 2.49-2.00	No	79	282	361
	Yes	61	233	294
	Total	140	515	655
☐ 2.74-2.50	No	40	119	159
	Yes	37	186	223
	Total	77	305	382
☐ 2.99-2.75	No	21	125	146
	Yes	32	195	227
	Total	53	320	373
☐ 3.49-3.00	No	44	219	263
	Yes	54	465	519
	Total	98	684	782
☐ 3.50-3.99	No	23	145	168
	Yes	38	521	559
	Total	61	666	727
☐ 4.00	No	2	26	28
	Yes	8	142	150
	Total	10	168	178
Grand Total		615	3,109	3,724



Data Warehouse Design Review

- Fact Tables
 - Occurrence tables of the Data Warehouse
 - Each row represents something happening in the organization
 - Facts either correspond to a measurable event (measure) or a relationship between enterprise entities (coverage tables)
 - Typically narrow tables with many rows

Fact Tables

<u>Student</u>	<u>School</u>	<u>Date</u>	<u>Sections Missed</u>	<u>Day Attended</u>
6094	33	20060601	1	1
6094	33	20060606	1	1
6094	33	20060607	2	0.5
6094	33	20060915	1	1
6094	33	20060918	8	0
6094	33	20060925	1	1
6094	33	20061003	1	1
6094	33	20061005	1	1
6094	33	20061019	1	1
6094	33	20061024	1	1

Coverage Tables

<u>dimstudent id</u>	<u>dimschool id</u>	<u>entry date</u>	<u>exit date</u>	<u>fteid</u>	<u>entry code</u>	<u>grade level</u>	<u>school enrollment days</u>
36091	39	20100825	20110612	1368	OE	0	176
36091	39	20110823	20120610	1969	NE	1	176

Data Warehouse Design Review

- Dimension Tables
 - Describe facts by linking them to a related set of attributes
 - Student
 - Employee
 - Date
 - Define the organization
 - Used as UI, Slicers and Rows (and usually columns) of reports

Dimensions

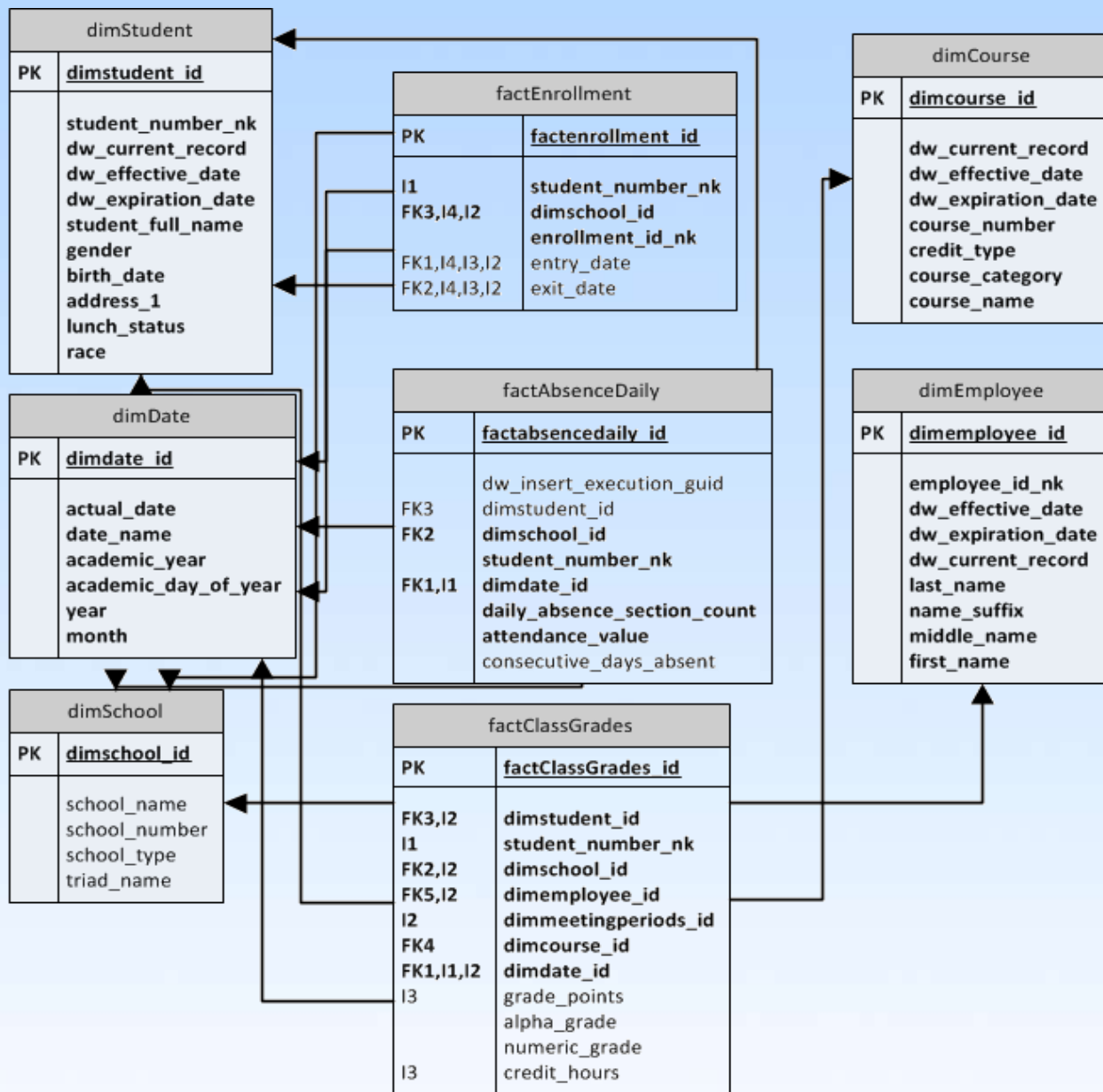
<u>student id</u>	<u>student number</u>	<u>current record</u>	<u>effective date</u>	<u>expiration date</u>	<u>student last name</u>	<u>student first name</u>	<u>address 1</u>
36091	123447	Y	8/23/2011	12/31/9999	Smith	Kaia	123 McVicker Street

Data Warehouse Design Review

- Type 2 Dimensions

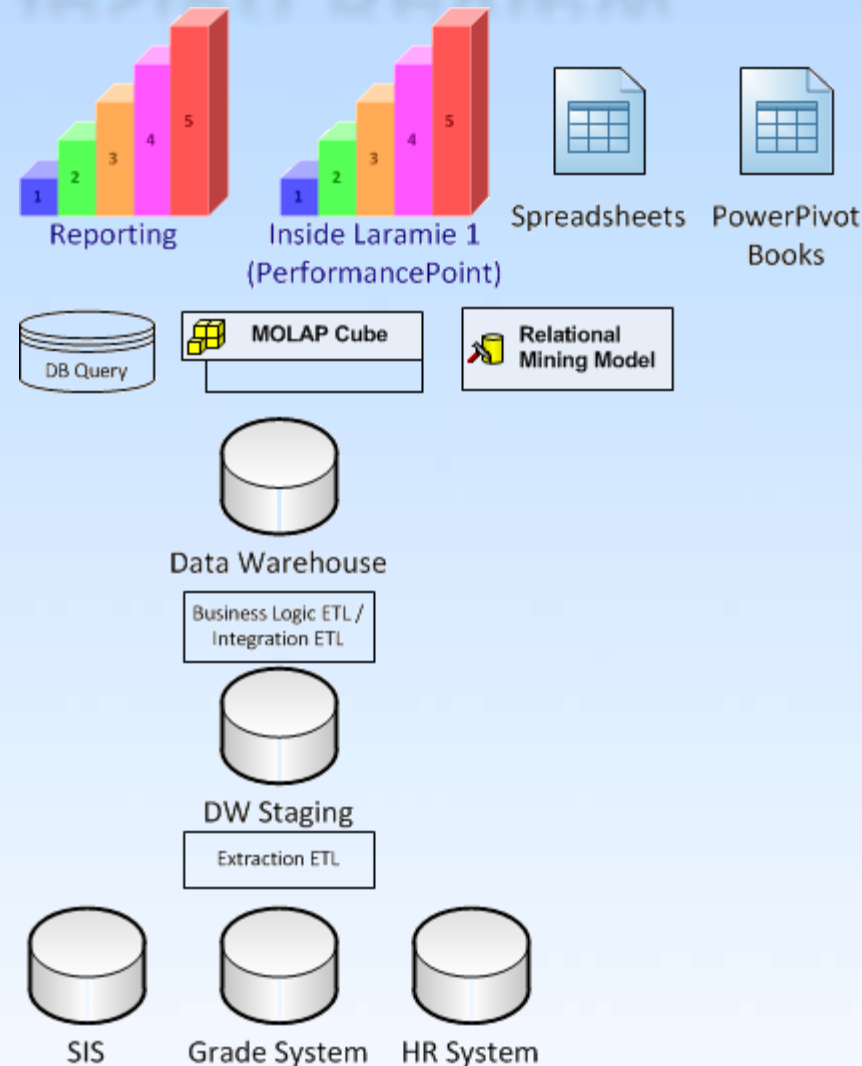
<u>student id</u>	<u>student number</u>	<u>current record</u>	<u>effective date</u>	<u>expiration date</u>	<u>student last name</u>	<u>student first name</u>	<u>address 1</u>
36091	123447	Y	8/23/2010	12/31/9999	Smith	Kaia	284 Center Avenue

<u>student id</u>	<u>student number</u>	<u>current record</u>	<u>effective date</u>	<u>expiration date</u>	<u>student last name</u>	<u>student first name</u>	<u>address 1</u>
36091	123447	N	8/23/2010	02/12/2011	Smith	Kaia	284 Center Avenue
456464	123447	Y	02/13/2011	12/31/9999	Smith	Kaia	123 3rd Street



Data Warehouse Design Review

- Enterprise Data Warehouse Structure
 - Extract Transform Load (ETL)
 - Data Warehouse Database
 - Operational Data Store
 - Data Cube
 - End User Interfaces
 - Reports
 - Analytical Tools
 - Data Extracts
 - Excel, PowerPivot, etc.



K-12 Dimensional Modeling Challenges

- Enrollment information is not a true student attribute
 - Opposed to demographics and other attributes describing just the student
 - Enrollment is more descriptive of a student's career and is usually much more volatile than other student attributes
 - But we like to talk about schools and grade levels, what do we do?

K-12 Dimensional Modeling Challenges

- Model enrollment information as a coverage fact table
 - Student ID, School ID, Grade Level, Begin and End Dates
- Solves granularity problem, but introduces new ones
 - Hard to query directly – have to pick a day to analyze
 - MDX does not handle between relationships well

K-12 Dimensional Modeling Challenges

- For cube relationships to function without complex code, we need a mechanism that:
 - Can link student, school and grade level information
 - Keeps track of date sensitive relationships
 - Example: attendance to proper school and grade level over time
 - Manages type 2 changes to student

Day Level Student Dimension Extension

- One row for every day of every student's career
 - Provides a “target” for everything a student could do
 - Fits back into student dimension's type 2 dates by selecting correct “version” of student based on which day of the career we are currently on
 - Carries foreign keys to student, school and date dimensions along with a grade level degenerate dimension

Enrollment and School Service Data

Student Enrollment Fact Table

<u>dimstudent id</u>	<u>dimschool id</u>	<u>entry date</u>	<u>exit date</u>	<u>fteid</u>	<u>entry code</u>	<u>grade level</u>	<u>school enrollment days</u>
36091	39	20100825	20110612	1368	OE	0	176
36091	39	20110823	20120610	1969	NE	1	176

School Service Fact Table

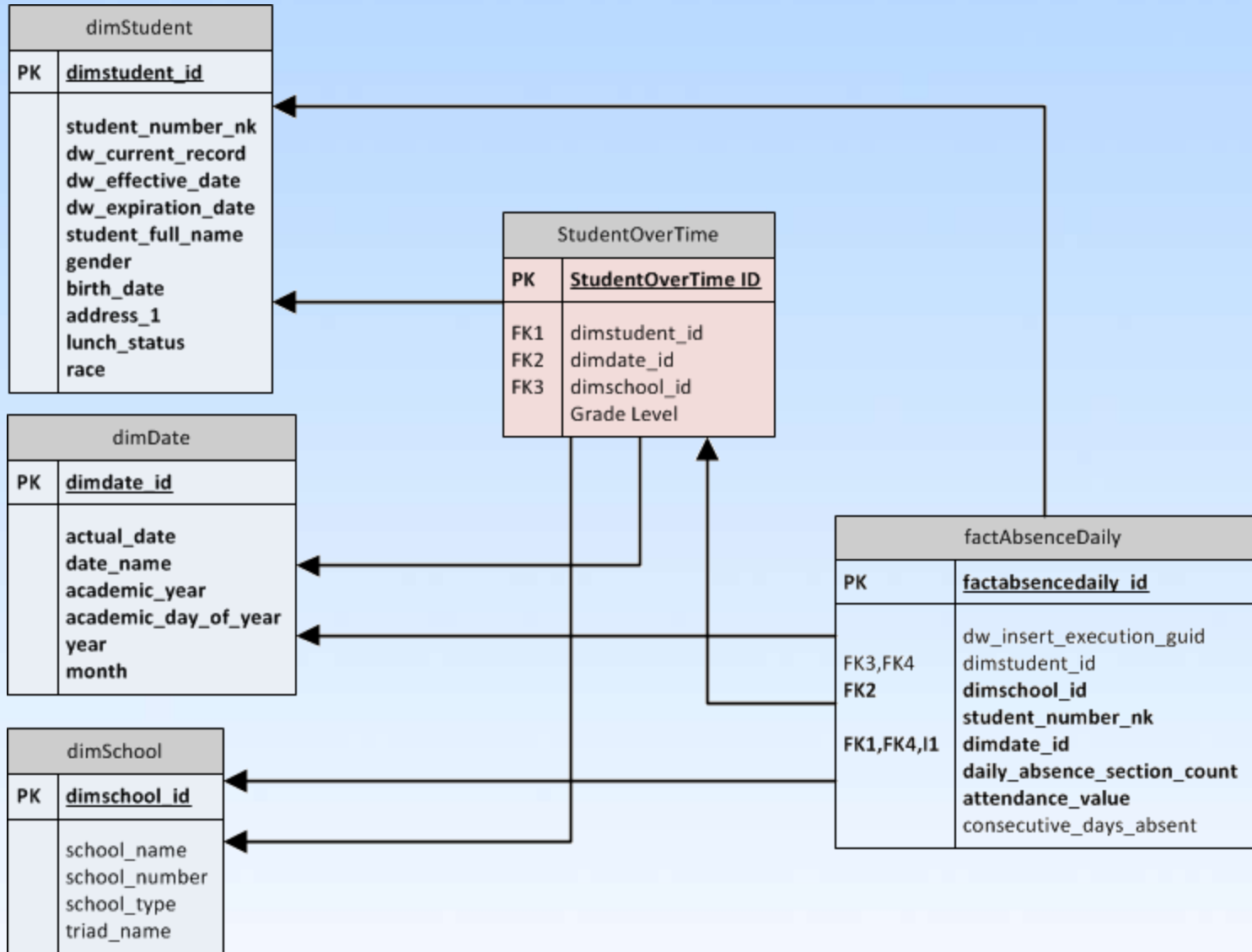
<u>dimdate id</u>	<u>dimschool id</u>	<u>abday</u>	<u>in session</u>	<u>membership value</u>
20110608	39	A	N	0
20110609	39	A	N	0
20110610	39	A	N	0
20110611	39	A	N	0
20110612	39	A	N	0
20110823	39	A	Y	1
20110824	39	A	Y	1
20110825	39	A	Y	1
20110826	39	A	Y	1
20110827	39	A	N	0
20110828	39	A	N	0
20110829	39	A	Y	1

Day Level Student Dimension Extension

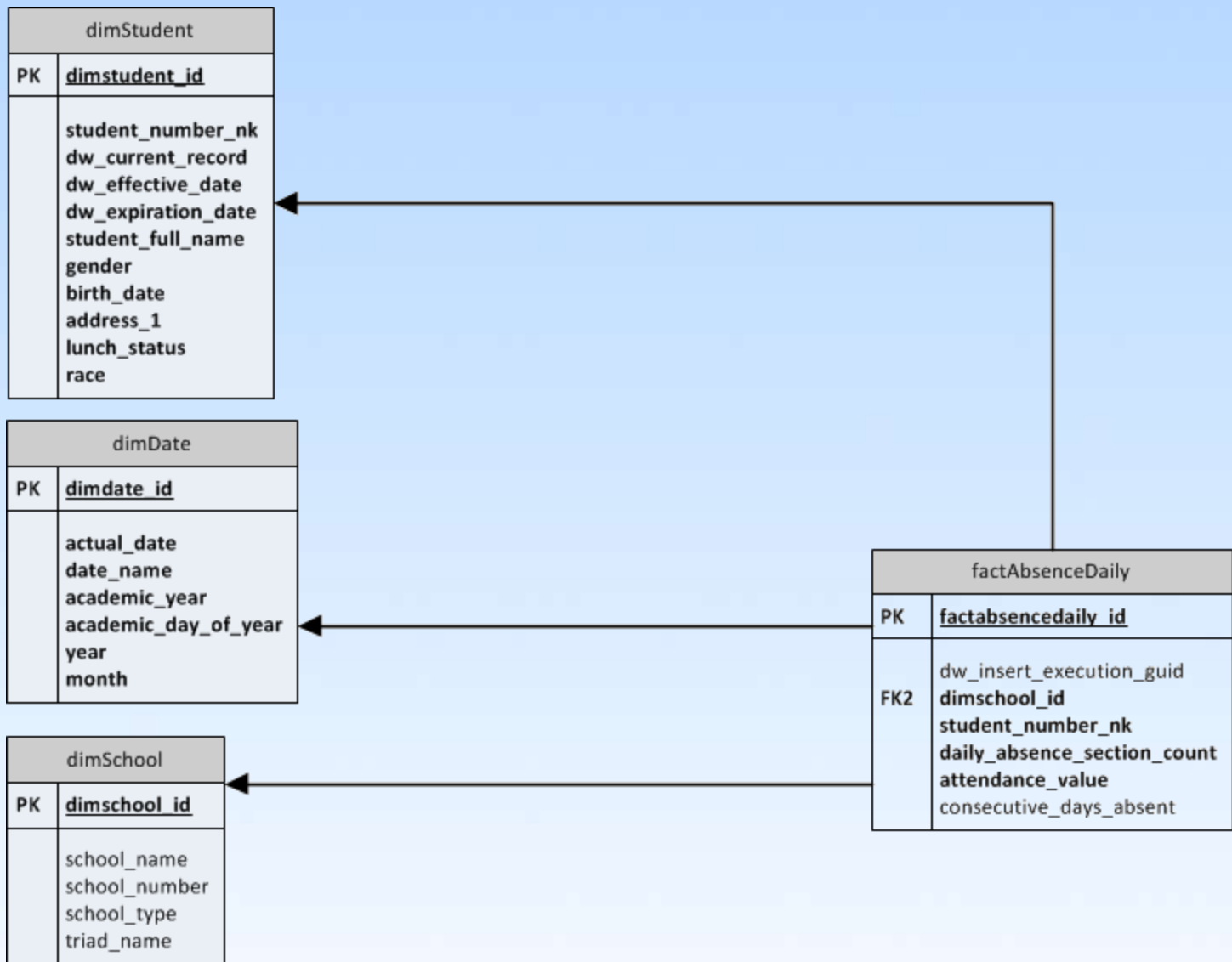
<u>vdimStudentOverTime id</u>	<u>dimstudent id</u>	<u>dimschool id</u>	<u>dimdate id</u>	<u>grade level</u>
45975624	36091	39	20110606	0
45975625	36091	39	20110607	0
45975626	36091	39	20110608	0
45975627	36091	39	20110609	0
45975628	36091	39	20110610	0
45975629	36091	39	20110611	0
45975630	36091	39	20110823	1
45975631	36091	39	20110824	1
45975632	36091	39	20110825	1
45975633	36091	39	20110826	1
45975634	36091	39	20110827	1
45975635	36091	39	20110828	1
45975636	36091	39	20110829	1

Day Level Student Dimension Extension

- Not directly changing the DW, sitting on top of it for reporting and cube



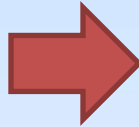
Note: All student info is fictitious



Comparison of Models

From base tables, each date needs to be specified

```
WITH reportdates
  AS ( SELECT 20110601 AS effdate
        UNION
        SELECT 20100601 AS effdate
      )
SELECT rd.effdate ,
       ds.race ,
       COUNT(*) AS count
FROM   reportdates rd
       CROSS JOIN dbo.dimStudent AS ds
       JOIN dbo.factEnrollment AS fe ON effdate >=
fe.entry_date
                                AND effdate < fe.exit_date
                                AND fe.dimstudent_id =
ds.dimstudent_id
GROUP BY rd.effdate ,
         ds.race ;
```



**From student over time, all dates are automatic;
we actually need to limit the dates**

```
SELECT fsot.dimdate_ID ,
       ds.race ,
       COUNT(*)
FROM   dbo.factStudentOverTime AS fsot
       JOIN dbo.dimStudent AS ds ON ds.dimstudent_id =
fsot.dimstudent_id
GROUP BY fsot.dimdate_ID ,
         ds.race ;
```

MDX and cube design is even more complex

So what did we build?

- A new dimension at a lower grain to the student table, snow-flaking the student table
- A rapidly changing dimension
- A mechanism that uses and tracks states of a student attribute over time
- The new bridge table does not change any of the underlying structures or relationships

So what did we build?

- Day-to-day grain allowed us to expand student descriptive attributes in more user-friendly ways from other fact tables:
 - GPA
 - Suspension and absence information
 - Counts to date
 - Consecutive streaks as of a day
 - Activity Enrollment
 - Assessment Status
- This would be unfeasible in a standard dimension

Derived Student Attributes

- Flexible model for changing or developing needs of the data model
 - Outside of the tightly coupled data model
- Does not have to be over time, current version would be very similar to a type 1 change

Student Attribute Driven Analysis

(Students within a population with an interesting attribute)

Students within a population

*Within a time frame or on a given day

```
SELECT
  --students who have done something interesting within a population
  SUM(CASE WHEN fsot.asOfGPA >= 3.5 THEN 1 ELSE 0 END) * 1.0
  /
  --students in that population
  COUNT(*)
  --population definition
FROM dbo.factStudentOverTime AS fsot
  WHERE fsot.membership_value = 1 AND fsot.dimdate_id = 20120201;
```

Basic Enrollment Count

Membership Value
3,724

Basic Enrollment Count

Membership Value
3,724



Slice by GPA



GPA Bucket ▼	Membership Value
.99-0.00	54
1.99-1.00	745
2.49-2.00	958
2.74-2.50	561
2.99-2.75	569
3.49-3.00	1,191
3.50-3.99	1,111
4.00	276
Grand Total	5,465

Basic Enrollment Count

Membership Value
3,724



Slice by GPA



GPA Bucket ▼	Membership Value
.99-0.00	54
1.99-1.00	745
2.49-2.00	958
2.74-2.50	561
2.99-2.75	569
3.49-3.00	1,191
3.50-3.99	1,111
4.00	276
Grand Total	5,465

Add columns breaking
Out a mobility attribute

	District 912 Contiguous ▼		
	No	Yes	Grand Total
GPA Bucket ▼	Membership Value	Membership Value	Membership Value
.99-0.00	16	32	48
1.99-1.00	160	419	579
2.49-2.00	140	515	655
2.74-2.50	77	305	382
2.99-2.75	53	320	373
3.49-3.00	98	684	782
3.50-3.99	61	666	727
4.00	10	168	178
Grand Total	615	3,109	3,724

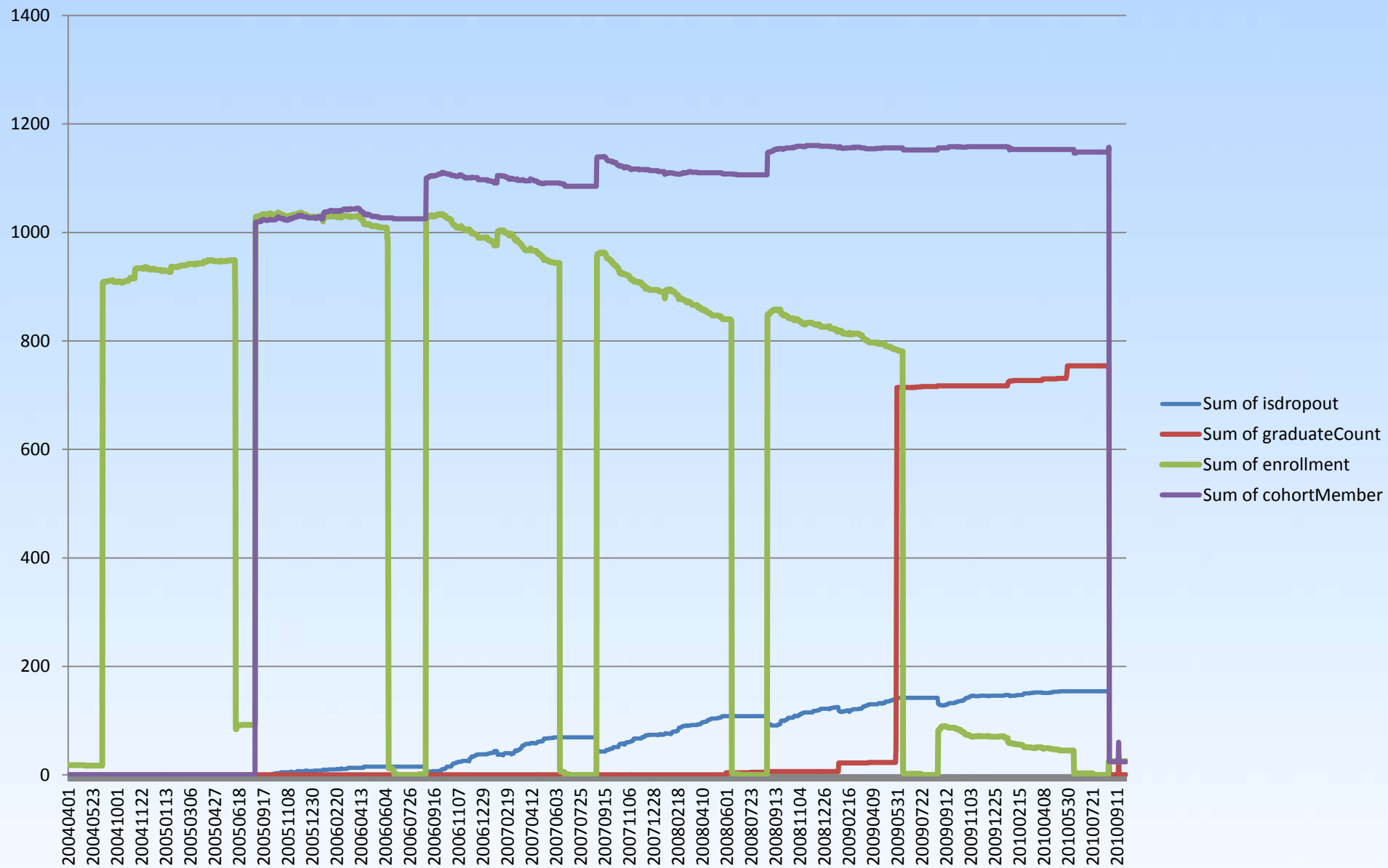
		District 912 Contiguous ▼		
		No	Yes	Grand Total
GPA Bucket ▼	Has Less Than Five Daily Absences YTD ▼	Membership Value	Membership Value	Membership Value
☐ .99-0.00	No	14	29	43
	Yes	2	3	5
	Total	16	32	48
☐ 1.99-1.00	No	106	304	410
	Yes	54	115	169
	Total	160	419	579
☐ 2.49-2.00	No	79	282	361
	Yes	61	233	294
	Total	140	515	655
☐ 2.74-2.50	No	40	119	159
	Yes	37	186	223
	Total	77	305	382
☐ 2.99-2.75	No	21	125	146
	Yes	32	195	227
	Total	53	320	373
☐ 3.49-3.00	No	44	219	263
	Yes	54	465	519
	Total	98	684	782
☐ 3.50-3.99	No	23	145	168
	Yes	38	521	559
	Total	61	666	727
☐ 4.00	No	2	26	28
	Yes	8	142	150
	Total	10	168	178
Grand Total		615	3,109	3,724

Transition to a fact-dimension

- Add a column with a count of one
 - School enrollment
 - Cohort tracking
 - Need additional rows added to support post “proper” graduation date rate tracking
- Additional attribute columns can be worked back into the student dimension
 - Class
 - Cohort Year
 - Career dropout and graduation status and dates

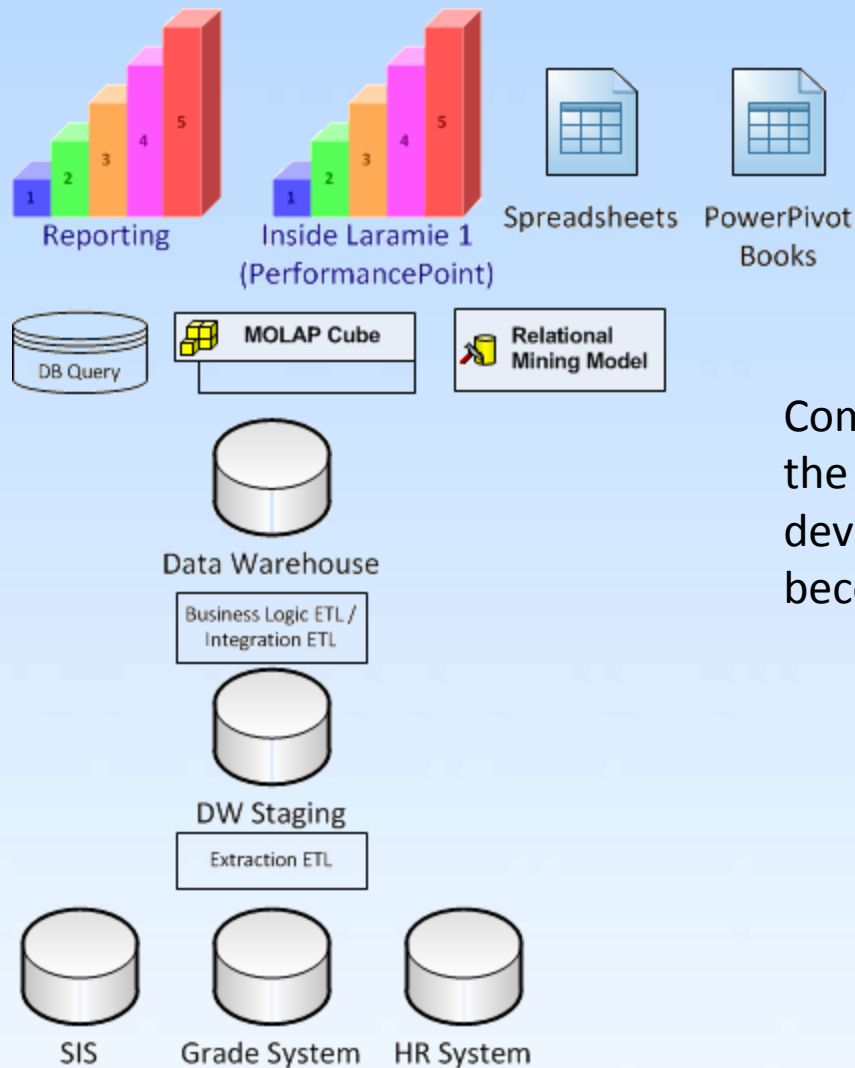
Transition to a fact-dimension

- Trending over time becomes trivial



Derived Attributes

- Business logic can be moved down the EDW stack to present hard to calculate metrics and attributes to end users
 - Mobility Status
 - Graduation Eligibility
- Slowly changing attributes can be worked back into the student dimension
- Rapidly changing attributes fit in the over time dimension



Complexity of deriving attributes moves from the top of the stack (user querying, custom development per report) to ETL layer and becomes an Enterprise asset

Q & A

Download presentation:

<http://goo.gl/7D1p5>

Contact Information

- Trevor Swarm
swarmt@laramie1.org
- Kyle McKinney
mckinneyk@laramie1.org
- Dr. Alan Moore
mooreal@laramie1.org

Additional Info (if time)

Building an over time fact-dimension

- Join a school service table with each student's enrollment to build “student days”
- Use an intermediate structure with student id, attribute value, begin and end date from base tables
- Most intermediate tables are generated with generic, reusable algorithms
- Join on student and the over time day being in range of the effective dates from the intermediate structure

Scalability Notes

- Views vs. materialized tables
 - Views - lightweight logic
 - Tables - the results of heavy business logic processing (intermediate effective date tables are a good first place)
- OK to scale back to only days you want to report on
 - Last two weeks of school days, end of each quarter

Scalability Notes

- Partitioning
- Change processing
- Indexing